BAYESIAN COMPUTATIONAL MODELS FOR INHARMONICITY IN MUSICAL INSTRUMENTS

*Simon Godsill**

Signal Processing Laboratory
University of Cambridge
Trumpington Street, Cambridge, CB21PZ, England
sjg@eng.cam.ac.uk

Manuel Davy

LAGIS/CNRS, Cité scientifique,
BP 48, 59651 Villeneuve d'Ascq, France
Manuel.Davy@ec-lille.fr

ABSTRACT

In this paper we describe recent advances in harmonic models for musical signal analysis. In particular we consider cases where inharmonic musical instruments are present. A particular model corresponding to a plucked string instrument is adopted and its parameters are automatically estimated from the data. In addition we describe new Bayesian prior distributions for musical pitch, which allow automatic adjustment to the pitch of the instruments, adaptive modelling of harmonic decay across frequency and automatic selection of the basis function window length. The methods are implemented using sophisticated MCMC algorithms and results demonstrated on a small set of examples, including harpsichord and guitar. The methods are readily adapted to polyphonic settings for pitch transcription and can be expected to lead to improvements in that setting.

1. INTRODUCTION

Computational music analysis is required in a number of audio applications, including estimation of musical instrument parameters for synthesis methods, extraction of features for classification of musical sounds, automated transcription, noise reduction and source separation. In computational music modelling it is required to provide accurate but simple models for musical audio, in order to facilitate these applications. One such model is the harmonic model, developed in Bayesian and non-Bayesian settings in [1, 2, 3, 4, 5] from origins in sin plus AR audio models, see [6]. See also [7] for related models within a matching pursuit setting. Recall that the basic model for a frame of audio data comprising a single note is as follows:

$$y_t = v_t + \sum_{m=1}^{M} \alpha_{m,t} \cos\left[\omega_m(B,\omega_0)t\right] + \beta_{m,t} \sin\left[\omega_m(B,\omega_0)t\right]$$

where t is the time index, M is the number of partials, ω_0 is the fundamental frequency, $\alpha_{m,t}$ and $\beta_{m,t}$ denote time varying amplitudes for the mth partial, and the term $\omega_m(B,\omega_0)$ represents the frequency of the mth partial. v_t is additive background noise, assumed to follow a Gaussian autoregressive process having parameters $\mathbf{a} = [a_1, ..., a_P]$, as in [4, 5]. In the purely harmonic case, $B = 0$ and we have

$$\omega_m(B,\omega_0) = m\omega_0$$

In [4, 5] inharmonic models were presented that allowed for small random deviations of frequency away from the 'ideal' frequency $m\omega_0$. This leads to quite a complex estimation procedure, since a random perturbation needs to be learned for each partial. In some cases, however, a more parsimonious model for the inharmonicity can be exploited based on the physics of the musical instruments. In this paper we consider unforced string oscillations. In particular for strings in unforced oscillation with finite stiffness the following is a well known approximate formula for the partial frequencies [8]:

$$\omega_m(B,\omega_0) = m\omega_0\sqrt{(1 + Bm^2)} \qquad (1)$$

where B is a physical parameter of the string. We will assume that B is unknown here and to be determined automatically from the data, along with all the other parameters in the above model. Such a model was also suggested, with fixed B, in [7]; estimation of B is a key component of our methods. Other physical models for inharmonicity could easily be incorporated within the framework by replacing this particular form for $\omega_m()$ with another parametric form. See also [9] and references therein for a full discussion of inharmonicity issues.

2. BAYESIAN MODEL

In this section we describe key aspects of the models, in particular where they depart from the previous work.

2.1. Partial amplitudes

As in [4, 5], partials are modelled with time-varying amplitudes $\alpha_{m,t}$ and $\beta_{m,t}$. This allows for the amplitude changes with time that are commonly observed in acoustic instruments. In such cases a standard Fourier analysis may be inappropriate since it implicitly assumes constant amplitude frequency components over the time frame of interest. The time-varying partials are parameterised in terms of the

* This work was partially funded by the EU Framework 5 project HASSIP and the EU NoE MUSCLE

0-7803-9154-3/05/$20.00

coefficients of some smooth basis functions ϕ_i, $i = 1, \ldots, I$ such that

$$\begin{aligned}\alpha_{m,t} &= \sum_{i=1}^{I} a_{m,i}\phi_{i,t} \\ \beta_{m,t} &= \sum_{i=1}^{I} b_{m,i}\phi_{i,t}\end{aligned}$$

There are many possible choices for the basis functions, and in practice any sufficiently smooth interpolation functions will do. As before we have implemented a simple scheme involving raised cosine functions (Hanning windows) with 50% overlap.

The basis coefficents $a_{m,i}$ and $b_{m,i}$ are assigned independent Gaussian prior distributions such that their amplitudes are assumed to decay with increasing frequency of the partial number m. The general form of this is

$$\begin{aligned}&p(a_{m,i}, b_{m,i}) \\ &\quad = p(a_{m,i})p(b_{m,i}) = \mathcal{N}(b_{m,i}|0, g\sigma_e^2 k_m)\mathcal{N}(a_{m,i}|0, g\sigma_e^2 k_m)\end{aligned}$$

Here g is a scaling factor common to all partials, σ_e^2 is the excitation variance of the residual autoregressive signal v_t, and k_m is a frequency-dependent scaling factor to allow for the expected decay with increasing frequency for partial amplitudes. As before, σ_e is an unknown quantity to be estimated by the MCMC algorithm. In a new extension of the models presented here, the amplitudes are supposed to decay as follows:

$$k_m = 1/(1 + (Tm)^\nu)$$

where ν is an unknown decay constant and T determines the cut-off frequency, see [10] for similar ideas within another setting. Such a model is based on empirical observations of the partial amplitudes in many real instrument recordings, and essentially just encodes a low pass filter with unknown cut-off frequency and decay rate. See for example the family of curves with $T = 5$, $\nu = 1, 2, ..., 10$, Fig. 1. It is worth pointing out that this model does not impose very stringent constraints on the precise amplitude of the partials: the Gaussian distribution will allow for significant departures from the $k_m = 1/(1 + (Tm)^\nu)$ rule, as dictated by the data, but it does impose a generally low-pass shape to the harmonics across frequency. It is possible to keep these parameters as unknowns in the MCMC scheme outlined below, and we have successfully experimented with this idea. In the simulations given later, however, T and ν were fixed to appropriate values for reasons of computational cost. g is treated as a random variable, assigned an inverted Gamma distribution for its prior with parameters $\alpha = 1.5$ and $\beta = 150$. This allows a good deal of variation from one note to another, while favouring partials that stand well above the autoregressive residual.

2.2. Basis function choice

Within the current framework it is feasible to choose automatically the number and type of basis functions $\phi_{i,t}$, see published discussion following [5]. For example, in the Hanning window case above with 50% overlap the choice would typically involve the time duration of each Hanning 'atom': shorter Hanning windows are better suited to non-stationary extracts with high degree of amplitude fluctuations, while longer Hanning windows are better suited to

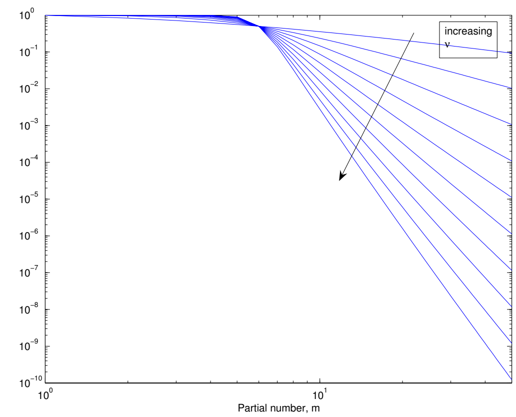

Figure 1: Family of k_m curves (log-log plot), $T = 5$, $\nu = 1, ..., 10$.

more stationary amplitude signals. Changes in window length will then implicitly change the number of atoms I in the representation above. We argue here that the window length should also be related to the fundamental frequency, since low frequency 'notes' can be expected to exhibit slower amplitude variation than higher frequency notes. In addition it is wise to prevent low notes from having high amplitude variability since this can lead to over-modelling of the data and amplitude/frequency ambiguities.

The specific proposal is as follows. The Hanning window length is specified as:

$$L(\omega_0) = \frac{32\pi\gamma}{\omega_0}$$

where $\omega_0 \in [0, \pi]$ is the normalised fundamental frequency and $L()$ is appropriate for a sampling rate of 44.1kHz. $\gamma \geq 1$ is then an unknown scaling factor that can scale the block length above the minimum length of $\frac{32\pi}{\omega_0}$. Here, for simplicity of implementation, we discretize γ to a small set of possible values, $\gamma \in \{1, 2, 4, 8\}$, and assign prior probabilities $\{0.1, 0.2, 0.3, 0.4\}$ to these possibilities. The results obtained are found to be fairly insensitive to these precise values, but the flexibility to vary the window length can be important.

2.3. Number of harmonics M

The number of partials is automatically determined from the data. The prior for M is uniform over some range $\{M_{\min}, ..., M_{\max}\}$, where $M_{\min} = 5$ and $M_{\max} = 100$ for the experiments reported below. In previous work [5] a simple single-stepping MCMC procedure was adopted for the sampling of M, in which proposals were randomly made to $M \pm 1$ and accepted using the Metropolis-Hastings (M-H) rule. This has proved too simple for the data observed below and we now adopt a scheme that draws a uniform random variable from the set $\Delta \in \{-5, ..., +5\}$ and then performs a M-H step proposing a change to $\min(M_{\max}, \max(M_{\min}, M + \Delta))$.

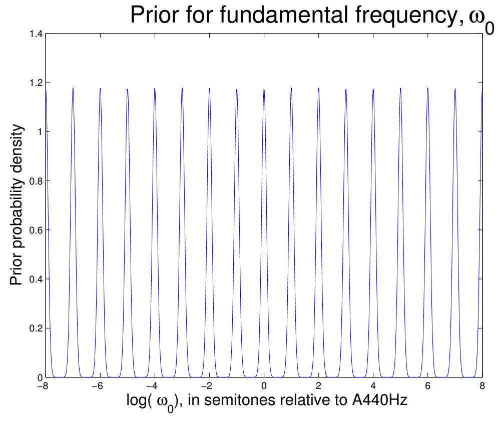

Figure 2: Prior for fundamental frequency

This has been found to explore the space of harmonic numbers much more effectively than before, allowing the algorithm to avoid various local maxima that could occur, such as when harmonic $M+1$ is very small or missing in the data, but $M+2$, ... are significant.

2.4. Fundamental frequency

Lastly, a note-based prior is used for the fundamental frequency ω_0. It is assumed that a Western tuning system applies and that notes lie at or close to the ideal pitches for that system. Assuming perfect tuning, these pitches are then $p_k = \Omega_0 2^{k/12}$, $k = k_{\min}...k_{\max}$, where Ω_O is the reference pitch for the system, e.g. A440Hz, and $k_{\min/\max}$ are the indexes of the lowest/highest note in the instruments.

The prior for ω_0, treated on a logarithmic scale is then

$$p(\log(\omega_0)) \propto f(\log(\omega_0) - \log(p_k)),$$
$$\log(\omega_0) \in (\log(p_k) - 1/24 \log 2, \log(p_k) + 1/24 \log 2]$$

where $f()$ is chosen as a narrow Gaussian shape in our work. See Fig. 2 for the shape used.

3. IMPLEMENTATION

Implementation is carried out using a Markov chain Monte Carlo (MCMC) algorithm [11]. This proceeds iteratively by proposing random changes to each of the parameters and accepting/rejecting these according to their posterior probabilities. Once converged, the scheme is used to obtain point estimates of the parameters, confidence intervals, and more generally the posterior distributions of any parameters of interest in the model. The precise algorithm is complex and cannot be detailed within the space confines of this paper.

4. RESULTS

A small selection of data has been processed in order to demonstrate the validity of the model and any improvements over previous purely harmonic models. We present the results graphically in the frequency domain, as this seems to give the clearest representation. Note, however, that the models and algorithms are all computed in the time domain.

Consider firstly a guitar pluck, which can be expected to exhibit inharmonicity, see e.g. [12]. A short section of length 4096 data points is extracted and analysed using the MCMC sampler[1]. Some outputs from the MCMC are shown in figure 3. In the top right panel the inharmonicity parameter B is seen to converge successfully to a plausible value within a few iterations. The estimated value is roughly $7 \times 10^{-5} \pm 1 \times 10^{-5}$. In the bottom left panel the input spectrum is compared with the estimated note spectrum. In the bottom right the residual spectrum is given, showing that the entire note has been successfully modelled with very little error. This was not the case when the equivalent model having no inharmonicity ($B = 0$) was implemented for the same data. Next a set of harpsichord examples was processed, see `http://www.acoustics.hut.fi/publications/papers/jasp-harpsy/` and [9]. For these examples inharmonicity values within a similar range were obtained, see for example the output from one note (file B3-8Bxx-28) in figure 5. Here we show in addition the MCMC exploration of the M parameter (number of partials) and the ω_0 parameter (almost immediate convergence to correct value for this parameter). Note that update rates were different in the sampler for each parameter - hence the differing total number of iterations listed. This harpsichord note has a much richer harmonic structure than the guitar note and hence a much more accurate estimate for the B parameter can be obtained. The results for five notes from this database are shown in the table, including the 5%/95%-iles obtained from the MCMC output.

Finally, studying the pitch 38 harpsichord note, we compare the use of the inharmonicity model with an exact harmonic model. This model is obtained by fixing $B = 0$ in the same MCMC code. Notice that there is a significant modelling mismatch in the upper partials when $B = 0$, which is not present when B is automatically estimated. In order to see this mismatch clearly, the minimum number of harmonics was set to be 22 in the harmonic case: otherwise the model does not introduce harmonics above around the 10th harmonic, as these are a poor match to the data. In the inharmonic case, the model is however quite capable of estimating the number of harmonics automatically. Notice how the inharmonic model fits the peaks in the spectrum with very high accuracy, while the harmonic model misses the high frequency peaks altogether. This result is an example of the improvement possible with this model, and also a validation of the approximate formula for ω_m given in Eq.(1). Misfitting of peaks as done by the harmonic model will inevitably lead to errors when the models are incorporated into a full transcription system, for example.

Pitch number	Mean	5%	95 %
04	2.273×10^{-5}	1.919×10^{-5}	2.56×10^{-5}
11	1.192×10^{-5}	1.19×10^{-5}	1.193×10^{-5}
28	3.175×10^{-5}	3.161×10^{-5}	3.192×10^{-5}
38	7.062×10^{-5}	6.988×10^{-5}	7.106×10^{-5}
56	39.49×10^{-5}	37.7×10^{-5}	42.6×10^{-5}

[1]Thanks to Jim Woodhouse for supplying this data

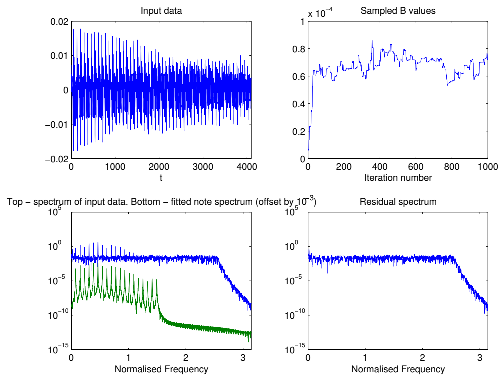

Figure 3: Results for single guitar pluck.

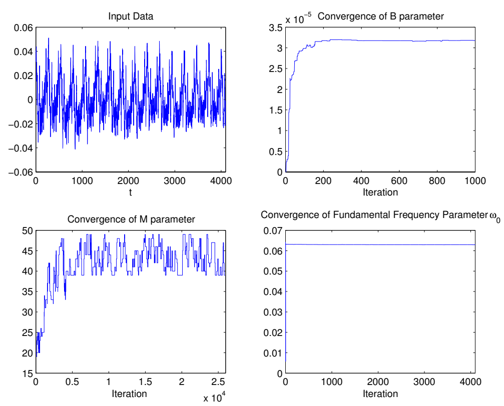

Figure 4: Results for single harpsichord note

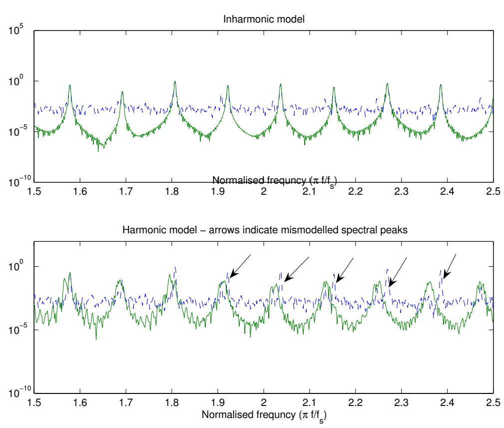

Figure 5: Results for single harpsichord note - harmonic vs inharmonic model. Dotted: true data, solid: fitted data

5. CONCLUSION

We have explored new versions of Bayesian harmonic models, adapted to inharmonic data. Effective estimates of note parameters have been obtained using MCMC, and good results are obtained using quite short data lengths. While it may be of interest to use these very accurate methods purely for analysis of inharmonicity in instruments, further work will incorporate these new models into automated transcription systems. This development should lead to improved recognition accuracy since individual notes will be better modelled. Currently we have experimented successfully with the new models in a multi-pitch setting, by extending to the multipitch case exactly as in our earlier work, although results are not yet extensive enough to report here.

6. REFERENCES

[1] R. Irizarry, "Local harmonic estimation in musical sound signals," *Journal of the American Statistical Association*, 2001.

[2] ——, "Weighted estimation of harmonic components in a musical sound signal," *Journal of Time Series Analysis*, vol. 23, pp. 29–48, 2002.

[3] P. J. Walmsley, S. J. Godsill, and P. J. W. Rayner, "Polyphonic pitch tracking using joint Bayesian estimation of multiple frame parameters," in *Proc. IEEE Workshop on Audio and Acoustics, Mohonk, NY State*, Mohonk, NY State, Oct. 1999.

[4] S. J. Godsill and M. Davy, "Bayesian harmonic models for musical pitch estimation and analysis," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2002.

[5] M.Davy and S. J. Godsill, "Bayesian harmonic models for musical signal analysis (with discussion)," in *Bayesian Statistics VII*, J. Bernardo, J. Berger, A. Dawid, and A. Smith, Eds. Oxford University Press, 2003.

[6] X. Serra, *Musical Signal Processing*. Swets and Zeitlinger, 1997, ch. Musical Sound Modeling With Sinusoids plus Noise.

[7] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Tr. Signal Proc.*, vol. 51, no. 1, 2003.

[8] N. Fletcher and T. Rossing, *The Physics of Musical Instruments*, 2nd ed. Berlin: Springer-Verlag, 1998, iSBN: 0-387-98374-0.

[9] J. K. M. L. Vesa Valimaki, Henri Penttinen and C. Erkut, "Sound synthesis of the harpsichord using a computationally efficient physical model," *Journal of Applied Signal Processing*, no. 7, pp. 934–948, 2004.

[10] S. Molla and B. Torrsani, "An hybrid audio scheme using hidden markov models of waveforms," *Applied and Computational Harmonic Analysis, to appear.*

[11] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds., *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall, 1996.

[12] J. Woodhouse, "On the synthesis of guitar plucks," *Acustica Acta Acustica*, vol. 90, pp. 579–589, 2004.